ChinaXiv合作期刊

一种卫星下传数据解析方法*

赵海升 樊凤霞 贾淑梅 李承奎
(中国科学院高能物理研究所北京 100049)

摘要：卫星采集的数据是按位或者字节的形式存储，这些格式都是地面预定义的。卫星下传的数据经过去格式及分包后，地面需要根据这些预定义格式将数据进行解析，生成可读的十进制数。这些格式尽管在不同卫星及载荷上有很大差异，但是它们都具有可解析性。本文设计了一种软硬件接口来定义及存储这些解析格式，同时根据这些格式产生软件可读的配置文件从而进行数据解析。这对大量数据文件来讲省时省力，同时可以保障解析的正确性。

关键词：卫星数据，数据包，数据解析，解析格式

中图分类号：p171.5

A decoding method of data packets for satellite data

Zhao Haisheng,Fan Fengxia, Jia Shumei, Li Chengkui

Key Laboratory of Particle Astrophysics, Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049, China

Abstract：The data packets obtained by satellite are stored as bits or bytesand their storage formats are defined by hardware engineers. When the data is telemetered to ground andthe same data packets will be organized into a binary format file. So, a large amount of files will need to be decoded into readable format files. Although their storage formats for different satellites or payloads have different styles, they can be decodedat same rule. This work will introduce a decoding system including an interface module and a decoding module. The interface module is used to store their storage formats to database and it makes a very convenient way to communicate between hardware engineers and software engineers. And a decoding module will be used to decode the data packets based on the data storage formats form database. This decoding method will save much time and labor cost.

Keywords：Satellite Data, Data Packets, Data Decoding, Data Format

0 引言

卫星采集的数据是按位或者字节的形式存储的，这些存储格式由地面预定义，下传数据的解析也都是遵照这些预定义格式进行。然而，卫星制造过程中，特别是其上载荷研制过程(硬件端)，

*基金项目：国家自然科学基金青年科学基金项目资助（11503029），中国科学院战略性先导科技专项资助，任务编号 XDA15310000.
作者：赵海升，男，博士，研究方向：天文数据分析及算法. E-mail：zhaohs@ihep.ac.cn

chinaXiv:201903.00227v1

这些格式时常发生变化，这就给数据解析（软件端）带来了很大麻烦；软硬件信息沟通，特别是沟通不及时，都会影响工程进展。另一方面即使这些格式固定，软件端解析这些量也费时费力。比较这些预定义格式，我们发现它们的定义规则比较一致，这样设计一种软硬件接口和数据解析方式，使之能够降低软硬件沟通成本，同时能够自适应的进行数据解析，就成为可能。

数据解析一般按数据包进行，卫星下传数据包有很多种，按采集来源可以分为卫星平台采集的数据包，载荷采集的工程数据包和载荷采集的科学数据包[1]。平台采集的数据包格式一般比较简单而且固定，比如定位数据包，姿态包及整星温度包等；载荷采集的工程数据一般是按时间采集的，数据种类比较多但是总的数据量小，比如反映驱动时钟状态的包，工作模式包等，一个包大约几十到几百个字节，有些时候硬件会在不同电子学或者仪器间轮换采集工程量，此时采集的工程量无论是格式，长度都可能不一样，它们可以看为单独的包，这样工程数据可能存在一些嵌套包；科学数据量大，考虑到星上存储资源比较珍贵，有些物理量分为两部分存储，有些情况下，一个物理事例（或量）的表征需要采集多个过阈值通道，每次过阈的通道数目不同，使得这种包长度不固定，比较重要的科学数据包有物理事例包，噪声基线包，时间进位包及 GPS 脉冲包等，一个包一般占用几个到几十个字节。这些原始数据包将在星上进行编码、加扰和调制，通过数传或者遥测通道下传至地面，地面再通过解调、解扰、解密，去格式，分包等，将同一类型的包按文件存储，这些不同的文件一般都有表征其唯一性的字段，比如硬 X 射线调制望远镜卫星（HXMT）[2]的文件名称采用 APID（应用过程标识符）表征。本文主要针对的即是这些去格式分包后的数据文件。

已经发射的 HXMT 和天宫二号上的 POLAR[3]探测器，星上数据存储形式是高位在前，低位在后，即对于一个字节的传输，首先传输高位，然后传输低位。无论是星上还是地面，数据流顺序都是按先传输，先存储，先读取的方式进行。在处理多字节数据时，地面人员需要注意字节顺序与计算机内存实际顺序的对应关系，某些计算机字节顺序会相反，比如本应该是字节 A 在字节 B 之前，但是内存中字节 B 在 A 之前，需要在解析时完成一次翻转即可。

本文主要介绍卫星下传数据的解析方法和软硬件接口方式，并将之纳入到一个系统中，希望方便各个卫星的调用，以节约软硬件交流成本，加速工程进度。

1 系统概述

本系统主要以数据库为中心，有两大模块。图 1 为系统框架，数据库的任务是存储解析格式，这些解析格式的形式将按本文给出的解析规则书写；模块一为接口模块，是硬件人员按照解析规则将解析格式写入到数据库，并能够再次读取数据库显示这些解析；模块二为数据解析模块，进行数据解析时，首先从数据库中读取解析格式，生成一系列配置文件，然后解析模块读取这些配置文件对数据进行解析。

chinaXiv:201903.00227v1

chinaXiv:201903.00227v1

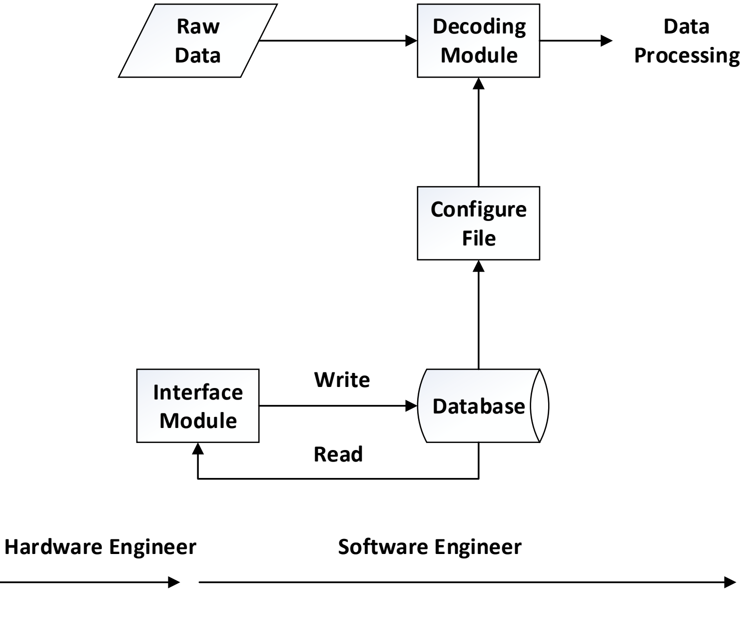


图 1 系统框架。

Fig. 1 Thearchitecture of decoding system.

数据库中存在两种类型的表格，一种表格为包性质表（注册表），它包含一类或者一系列包的属性定义；另一种为包格式表（格式表），每个包都对应一个格式表，它指明包的解析格式。这样在解析一种数据包的时候，首先从注册表中找到该包的属性，包括长度，起止关键字等，然后查询出需要的格式表从而找出解析格式。配置文件的内容与相对应的数据库中表格类似，名称亦相同，可以看为数据库的一个副本，故以下描述中统称为数据库。

1.1 注册表介绍

注册表表征包的一些基本属性，它包括如下信息：

title：包名称，对应格式表名称，具有唯一性；

type：包类型，这是识别包的关键信息。某一数据文件对应一个或者多个包，将它们的类型设为相同的，可以快速得到该文件对应的包。比如对于 HXMT，物理事例包，标定事例包的类型都是事例型。本文建议利用 APID 代表包类型；

header：表征数据包起始点，它由起始位置，结束位置和值组成，其中值可以有两个，这是因为有些载荷通过轮换 header 表征丢包信息，比如前一个包值为 0x103B，而后一个则为 0x183B；

mode：表征数据包性质，相同类型的包可能性质不同，比如 header 表征事例包，而 mode 进一步表征该包为科学事例包或者为标定事例包。它由起始位置，结束位置和值组成；

main：是否为主包，对于工程数据，一个文件一般对应一种主包，但是该包可以包含多个嵌套包；科学数据文件可能包含多种并列包，但是它们出现频次是不同的，对于频次最高的包定义为主包，以方便高频次调用；同类型包只能存在一种主包；

length：它由起始位置，结束位置及预设值表征，如果位置存在，则位置所在的值为包长度，否则查看预设长度，如果预设长度存在，则该类型的包长度固定，否则为可变长度包；

tail：表征结尾，由起始位置，结束位置及预设值表征，如果包为可变长度包，起始及结束位置应该不存在，建议硬件端设置预设值，比如 0xFFFF 表征包结束；

validation：特殊位置及其值检查，一般设置两处（起始位置，结束位置和值）可供选择；

crc[4]：包校验位置和长度，对于可变长度包，建议硬件端将位置设在 tail 之前。

为了方便解析，length 与 tail 必须存在其一，length 可以只有预设值。嵌套情况中（硬件设计），嵌套包 header 和 mode 必须存在其一，且为整字节，硬件设计的嵌套不是简单的将一部分解析打包，而是与电子学读出有关系，比如采集不同的电子学通道时，如果数据结构相同而采用嵌套或者采集的数据结构不同但数据量相同而采用嵌套。与之对应的是软件定义的嵌套，软件人员可以将一些数据结构定义为嵌套包，此时嵌套包不需要 header/mode，只给出唯一的 title 即可。另外主包及嵌套包如果存在 tail 时，需要设置不同的值以区分两者。非定长结构尽可能靠近包尾，并且尽量使一个包中仅有一个非定长结构。

其它的一些重要信息为：type 和 header/mode 确定唯一的格式表，title 唯一，且用于寻找格式表，validation 和 crc 用于做检查，一般而言如果 crc 校验失败，则该包会被丢弃。

比较常见的包一般定义 title，type，length，crc 即可，而 header，mode，tail，validation 都可以不存在。如果不存在，硬件端可以将其属性值均设为-1，比如 mode 不存在，则需要将其起始、结束位置及值均设为-1。起止位置为相对于包开始的位数或者字节数。

1.2 格式表介绍

格式表用于记录每种数据包的解析格式，它对应注册表的一个记录。它包括如下信息：

id：待解析变量的序列，自 0 开始，用户可以通过 id 获得解析量；

title：待解析变量名称，用户可以通过 title 获得解析量；

position：待解析变量的地址，可以存在两个位置，每个位置有起始位置，结束位置及标识表征，第一个位置为低位，后一个为高位。多数待解析变量只用一个位置表示，另一个位置的量可以设为-1；如果位置中仅有起始位置，而结束位置设为-1，则表示待解析量为不定长度；标识表征位置为绝对位置还是相对位置，前者表示自包开始的位置，而后者表示相对于上一解析量（id-1）的位置，对于不定长数据结构，其末尾可以占据整数字节的任意位，相对位置不固定，这样我们要求非定长结构需要占用整数字节，如不够则对剩余位进行填补；

type：待解析变量类型，0 表示一般变量或者连续型定长数组；1 表示连续型非定长数组，所谓非定长数组是指数组长度不定，需通过解析其它信息等获取；2 表示间断型非定长数组，此数组在内存上不连续，而是相邻数组元素间隔一定的长度，一般几个数组交替出现，这种结构对应多通道或者多单体数据采集。对于 type=0，position 表示整个解析区间；而对于 type=1，2，position 只表示数组第一个元素的位置；如果 type 为 3，表示待解析为包类型（嵌套包）；

reference：当 type 为 3 时使用，类型为字符串，其值为包名称（注册表中的 title），可以通过该值获取嵌套包解析格式；如果此处为空，表示该处可以放置多种嵌套包；

chinaXiv:201903.00227v1

repeat：当 type=0 时，此值表示数组维数（数组为连续的，将 position 指定的区间分为 repeat 份），对于非数组型，此值可以设为 1；当 type=2 时，repeat 表示间隔距离，比如第一个元素的位置为[a，b]，下一个元素的位置为[b+repeat+1，b+repeat+1+b-a]；

flag：表征解析单位为字节（BYTE）还是位（BIT）。

本文主要提供了两种可变长度，一种是连续性数组，另一种则为间断型数组，特别是后者，它实际可以用于多通道/多单体数据采集。为了解析方便，我们规定定义的非定长包只能有一种非定长结构。非定长结构只能占用整数字节，不足整个字节则填充预设值。下一个变量相对于非定长结构的位置也以字节结束为起始。

软件定义的嵌套包可以不按整数字节定义，但是需要有明确的引用，比如可以将非定长结构“ABCABCABC…..”中的“ABC”看为一个嵌套包，该结构就成为一个非定长数组结构。

2 接口模块

接口模块的主要任务是完成注册表和格式表的记录和显示，并生成解析文件。这一接口允许硬件端随意改变解析格式，而不影响软件端的解析。接口的实现采用基于 web 的设计和开发，以网站的形式向硬件和软件人员提供（或者采用 QT[5]也是不错的选择）。表格 1 为网站所涉及的人员及其权限。网站所涉及的层次结构如图 2 所示，一个卫星项目下主要分为平台包和载荷包，载荷包下可以分为很多类型，比如分为科学包和工程包，每种类别（本文看为数据类别）均对应一个注册表和若干格式表。

表格 1 接口模块涉及人员权限。

Table 1 The configuration of rights management.

人员	管理员	硬件人员	软件人员
权限	对软、硬件人员开通权限，限制权限；添加/修改/删除卫星	在其权限下：添加/修改/删除载荷；添加/修改/删除数据类别；完成格式表、注册表的填写/修改/删除；查看	在其权限下：查看注册表，格式表；生成解析文件

网站后台的数据库采用 MYSQL[6]，一个卫星项目可以对应其中的一个数据库，注册表和格式表对应其中的表格。数据库名称可以定义为卫星名称，注册表的命名构成为：卫星名称+“_”+载荷名称+“_”+数据类别名称，其中数据类别包括科学数据，工程数据等；格式表命名构成为：卫星名称+“_”+载荷名称+“_”+数据类别名称+包名称。

解析模块读取的是配置文件，而不是数据库，这样做的好处在于将解析模块和数据库解绑，两者不直接相关；另一方面，利用配置文件进行版本控制，载荷格式更新后，可以保存配置文件并记录时间，这在数据追溯中比较重要。

chinaXiv:201903.00227v1

chinaXiv:201903.00227v1

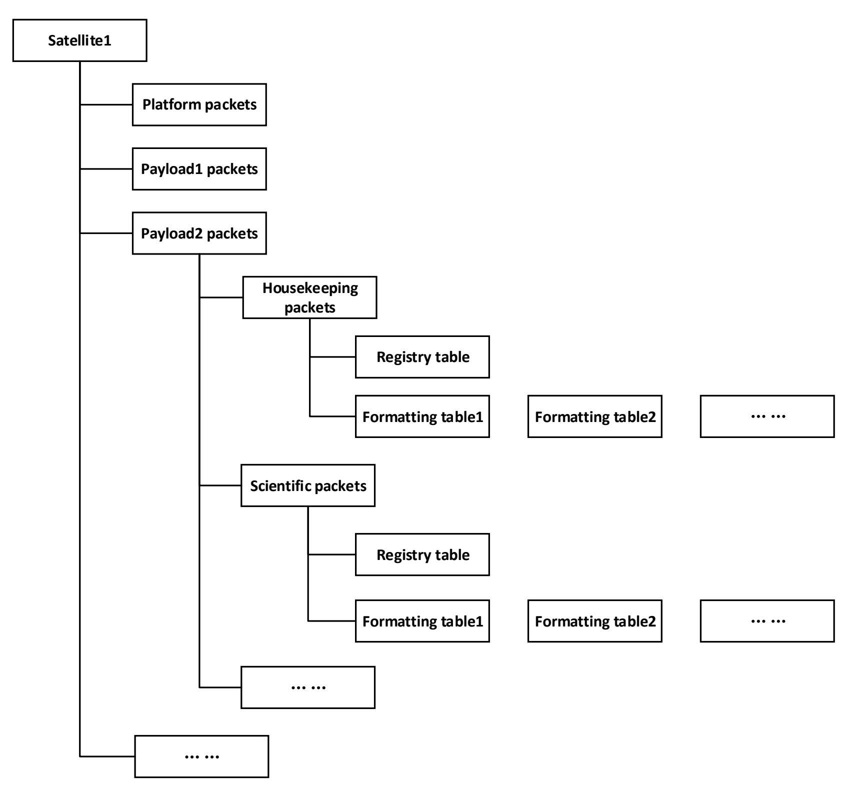


图 2 网站的层次结构。

Fig. 2 The framework of web sites.

3 解析模块

数据解析主要针对载荷测试数据和卫星下传数据，卫星下传数据比较有组织性，但是涉及多个单位和部门，需要协调接口，而前者组织比较松散，主要与载荷有关。本结将重点介绍文件的识别，注册表的读取，格式表的解析及对特殊数据的解析处理。

卫星下传数据一般都有表征数据类型的字段或者值，比如 HXMT 采用 APID 标记数据文件，它可以对应本文中的包类型（注册表中的 type）。我们建议下传数据文件和测试数据文件的文件命名规则包含 APID 或者是其它表征包类型的关键字。如果输入文件采用 APID 命名的话，需要建立 APID 与包类型，数据类别，载荷的对应关系。当然更简单的方式是名称中有包类别，数据类别和载荷关键字。

文件解析时读入文件名称，解析出其包类别，数据类别及载荷等信息。确定查询的注册表（注册表中 type 与包类型相同的行），并得到需要解析的格式表（注册表中的 title 对应格式表名称），需要注意的是一种数据文件可能对应多个格式表。

数据的获取，首先确定注册表查询结果中的主包（注册表中 main 标志），如果没有主包，则各个查询结果为并列结构。所谓并列结构，是指查询的各个格式表都可能出现，需要利用注册表中的关键字进行匹配，匹配成功者为这段数据需要的格式表。如果存在主包，首先进行关键字匹配，如果匹配成功，则调用该包的格式表，如果不成功（一般对应出现频次较高者人为设置为主包的情况），则需要对其它查询结果进行匹配。确定格式包后，可以完整的获取包信息，对于定

chinaXiv:201903.00227v1

长则取 length 关键字（注册表）指定的位/字节/字数目，对于非定长则搜寻数据的 tail 关键字。

关键字匹配中需要考虑 header，mode 和 validation 等关键字，对于 crc 校验不通过者可以将该包剔除（如果确定不了包长度，可以通过在数据中搜索 header，tail 等实现）。

获取解析结果需要明确非定长结构的长度，即数组维数，一般情况下该包的数据长度（length 和 tail 决定）已知，则可以计算非定长结构维数（非定长起点已知，且占整数字节，tail 相对于非定长结束的字节亦知，故可计算维数），更一般的情况是非定长结构与某些变量关联，即某变量的取值或者其位数（比如为 1 的数目）表征非定长结构的长度，如果用户已知其对应关系，则会非常简单的对非定长数据进行处理。

本系统的驱动可以是文件驱动型，即一旦检测到新文件传输完成，本系统自动进行文件的解析工作，但是需要用户预先定义好输出数据结构，数据获取接口及输出格式。本系统不提供“文件-文件”式的解析方式，这是因为这些解析量不需要全部记录，而只记录用户感兴趣的；而且文件中可能有多种数据解析格式和多种嵌套方式，不容易自动完成输出格式的编排；有些时候一个原始文件还需要产生多个其它文件。而且这样做弥补一些设计缺陷，比如对于变量之间的关联（某变量与非定长结构的长度关联），用户容易得到这些关系。数据获取主要是用户通过格式表中的 title 或者 id 得到解析后的变量，输出格式一般是 FITS[7]格式，需要注意的是对于一个解析文件，可能生成多个输出文件。

4 总结

数据解析系统是卫星数据产品生产中最先进行的一环，不同的卫星及载荷尽管在数据设计上不同，但是都遵循数据的可解析性，这就是本系统的依据。数据解析在地面测试中同样重要，而且随着载荷的推进，解析格式也会发生变化，本系统可以适应这些变化。本系统解析结果的正确性是保障获得物理成果的基础，文献[1]中提到的通用数据解析方法(2.1 节)实际上是本系统解析模块的一个简化版本，它成功的应用于 POLAR 和 ME[8]探测器地面数据解析中。本系统采用通用的解析方法，避免了软件人员直接处理大量的数据文件，这也是保障获得物理成果的一个重要因素。

卫星数据按包解析，解析后的数据处理和存储，本系统不涉及，而是需要用户自行定义存储数据的结构和存储格式，用户获取解析后的数据则是通过本系统的接口实现。当这些条件具备时，本系统即可自动运行并与后续的软件共同协作完成数据产品的产生。

参考文献

[1]赵海升，葛明玉，李正恒，聂建胤，宋黎明. 一种天文卫星数据预处理方法[J]. 天文研究与技术, 2017, 14(3): 376-381.

Zhao Haisheng, Ge Mingyu, Li Zhengheng, Nie Jianyin, Song Liming. A Method of Data

Preprocessing for Astronomical Satellite[J]. Astronomical Research and Technology,2017, 14(3): 376-381.

[2] Li Ti-Pei, Wu Mei. The hard X-ray modulation telescope mission[J]. Physics, 2008,37(09):648-651.

[3] N.Produit, Barao F, Deluit S, et al. POLAR:a compact detector for gamma-ray bursts photon polarization measurements[C], Nucl.Instr. and Meth.A, 2005, 550(3) : 616-625.

[4] Han Jiawei, Zhang Hongqun. Error correction technology based on CCSDS standard for remote sensing satellite[J]. Chin. J. Space Sci., 2009, 29(1):112-116.

韩家玮,张洪群，基于CCSDS标准的遥感卫星下行数据纠错校验技术[J]，空间科学学报, 2009, 29, Vol.3 (1):112-116.

[5] 周鑫磊，王威，王锋，邓辉，柳翠寅，胡杰，李少良，王璐璐，周士然. 基于 QT 的 MUSER 观测数据多屏图形化实时显示的设计与实现[J]. 天文研究与技术, 2015, 12(4): 503-509.

Zhou Xinlei, Wang Wei, Wang Feng, Deng Hui, Liu Cuiyin, Hu Jie, Li Shaoliang, Wang Lulu, Zhou Shiran. Design and Implementation of a Multi-Monitor Display System Based on the Qt for NAOC MUSER Observations[J]. ASTRONOMICAL RESEARCH & TECHNOLOGY, 2015, 12(4): 503-509.

[6] 李会贤，桑健，王沙，罗阿理，基于MYSQL/LUNUX的LAMOST数据库设计与实现[J],天文研究与技术, 2006, Vol.3 Issue(1):56-63.

LI Hui-xian, SANG Jian, WANG Sha, LUO A-l, The Database Design of LAMOST Based on MYSQL/LINUX, Astronomical Research and Technology[J], 2006, Vol.3 Issue(1):56-63

[7] Wells D C, Greisen E W & Harten R H. FITS: A Flexible Image Transport System[J]. A&AS, 1981, 44: 363-370; A Primer on the FITS Data Format[R/OL]: http://fits.gsfc.nasa.gov/fits primer.html.

[8] 曹学蕾,姜维春,张万昌,孟斌,杨生,雒涛,顾煜栋,谭颖.硬 X 射线调制望远镜卫星中能望远镜设计与验证[J].航天器工程,2018,27(05):127-133.

Cao Xuelei, Jiang Weichun, Zhang wanchang, Meng Bin, Yang Sheng, Luo Tao, Gu Yudong, Tan Ying, Design and verification of Medium Energy Telescope on board HXMT satellite[J], Space Craft Engineering, 2018,27(05):127-133.

chinaXiv:201903.00227v1